

Recognising Agreement and Disagreement between Stances with Reason Comparing Networks

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks

CSIRO Data61, Australia

{Chang.Xu, Cecile.Paris, Surya.Nepal, Ross.Sparks}@data61.csiro.au

Abstract

We identify agreement and disagreement between utterances that express stances towards a topic of discussion. Existing methods focus mainly on conversational settings, where dialogic features are used for (dis)agreement inference. We extend this scope and seek to detect stance (dis)agreement in a broader setting, where *independent* stance-bearing utterances, which prevail in many stance corpora and real-world scenarios, are compared. To cope with such non-dialogic utterances, we find that the reasons uttered to back up a specific stance can help predict stance (dis)agreements. We propose a reason comparing network (RCN) to leverage reason information for stance comparison. Empirical results on a well-known stance corpus show that our method can discover useful reason information, enabling it to outperform several baselines in stance (dis)agreement detection.

1 Introduction

Agreement and disagreement naturally arise when peoples' views, or "stances", on the same topics are exchanged. Being able to identify the convergence and divergence of stances is valuable to various downstream applications, such as discovering subgroups in a discussion (Hassan et al., 2012; Abu-Jbara et al., 2012), improving recognition of argumentative structure (Lippi and Torroni, 2016), and bootstrapping stance classification with (dis)agreement side information (Sridhar et al., 2014; Ebrahimi et al., 2016).

Previous efforts on (dis)agreement detection are confined to the scenario of natural dialogues (Misra and Walker, 2013; Wang and Cardie, 2014; Sridhar et al., 2015; Rosenthal and McKeown, 2015), where dialogic structures are used to create a conversational context for (dis)agreement inference. However, non-dialogic stance-bearing utterances are also very

common in real-world scenarios. For example, in social media, people can express stances autonomously, without the intention of initiating a discussion (Mohammad et al., 2016). There are also corpora built with articles containing many self-contained stance-bearing utterances (Ferreira and Vlachos, 2016; Bar-Haim et al., 2017).

Studying how to detect (dis)agreement between such *independent* stance-bearing utterances has several benefits: 1) pairing these utterances can lead to a larger (dis)agreement corpus for training a potentially richer model for (dis)agreement detection; 2) the obtained pairs enable training a distance-based model for opinion clustering and subgroup mining; 3) it is applicable to the aforementioned non-dialogic stance corpora; and 4) it encourages discovering useful signals for (dis)agreement detection beyond dialogic features (e.g., the reason information studied in this work).

In this work, we investigate how to detect (dis)agreement between a given pair of (presumably unrelated) stance-bearing utterances. Table 1 shows an example where a decision is made on whether two utterances agree or disagree on a discussed topic. This task, however, is more challenging, as the inference has to be made without using any contextual information (e.g., dialogic structures). To address this issue, one may need to seek clues *within* each of the compared utterances to construct appropriate contexts.

Topic: Gun Control	
Utterance 1: <i>If guns are outlawed, only outlaws will have guns.</i> (Stance: Against)	Utterance2: <i>Freedom to have a gun is same as freedom of speech.</i> (Stance: Against)
Class Label: Agree	

Table 1: The task of detecting stance (dis)agreement between utterances towards a topic of discussion.

It has been observed that when expressing

stances, people usually back up their stances with specific explanations or reasons (Hasan and Ng, 2014; Boltuzic and Snajder, 2014). These reasons are informative about which stance is taken, because they give more details on how a stance is developed. However, simply comparing the reasons may not be sufficient to predict stance (dis)agreement, as sometimes people can take the same stance but give different reasons (e.g., the points *outlaws having guns* and *freedom of speech* mentioned in Table 1). One way to address this problem is to make the reasons *stance-comparable*, so that the reason comparison results can be stance-predictive.

In this paper, in order to leverage reason information for detecting stance (dis)agreement, we propose a *reason embedding* approach, where the reasons are made stance-comparable by projecting them into a shared, embedded space. In this space, “stance-agreed” reasons are close while “stance-disagreed” ones are distant. For instance, the reason points *outlaws having guns* and *freedom of speech* in Table 1 would be near to each other in that space, as they are “agreed” on the same stance. We learn such reason embedding by comparing the reasons using utterance-level (dis)agreement supervision, so that reasons supporting agreed (disagreed) stances would have similar (different) representations. A reason comparing network (RCN) is designed to learn the reason embedding and predict stance (dis)agreement based on the embedded reasons. Our method complements existing dialogic-based approaches by providing the embedded reasons as extra features. We evaluate our method on a well-known stance corpus and show that it successfully aligns reasons with (dis)agreement signals and achieves state-of-the-art results in stance (dis)agreement detection.

2 RCN: The Proposed Model

Figure 1 illustrates the architecture of RCN. At a high level, RCN is a classification model that takes as input an utterance pair (P, Q) and a topic T , and outputs a probability distribution \mathbf{y} over three classes $\{Agree, Disagree, Neither\}$ for stance comparison. To embed reasons, RCN uses two identical sub-networks (each contains an RNN encoder and a reason encoder) with shared weights to extract reason information from the paired utterances and predict their stance (dis)agreement based on the reasons.

RNN Encoder: In this module, we use RNNs to encode the input utterances. We first use *word embedding* to vectorise each word in the input utterance pair (P, Q) and topic T , obtaining three sequences of word vectors \mathbf{P} , \mathbf{Q} , and \mathbf{T} . Then we use two BiLSTMs to encode the utterance and topic sequences, respectively. Moreover, by following the work of Augenstein et al. (2016), we use *conditional encoding* to capture the utterances’ dependencies on the topic. The output are two topic-encoded sequences produced by the utterance BiLSTM for $P(Q)$, denoted by $\mathbf{H}^{P(Q)} = \{\mathbf{h}_i^{P(Q)}\}_{i=1}^{|P(Q)|} \in \mathbb{R}^{|P(Q)| \times 2h}$, where h is the hidden size of a unidirectional LSTM.

Reason Encoder: Then we extract reasons from the utterances, which is the main contribution of this work. In particular, we focus on the *major* reasons that most people are concerned with, which possess two properties: 1) they are focal points mentioned to support a specific stance; 2) they recur in multiple utterances. With such properties, the extraction of these reasons can then be reduced to finding the *recurring focal points* in all the input utterances.

To action on this insight, we take a weighting-based approach by learning a weighting matrix \mathbf{A} that captures the relatedness between each position in an utterance and each implied reason. For example, on utterance P where we hypothesise κ possible reasons, the weighting matrix is $\mathbf{A}^{|P| \times \kappa}$, with each cell $\mathbf{A}_{i,k}$ representing the relatedness between the i th position of P and the k th reason.

To learn the weighting matrix \mathbf{A} , we use *self-attention* (Cheng et al., 2016) and develop a particular self-attention layer for implementing the above weighting scheme. Meanwhile, the recurrence of a reason is also perceivable, as all utterances mentioning that reason are used to learn the

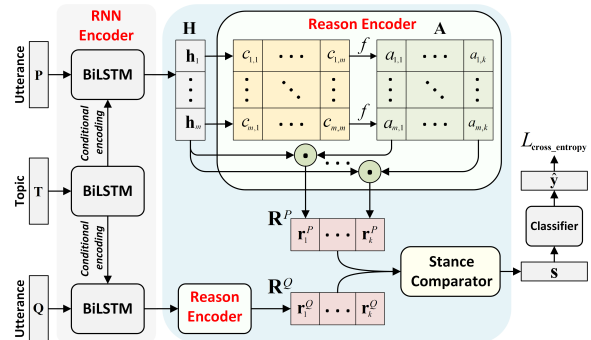


Figure 1: The architecture of RCN.

self-attention layer.

Our particular self-attention applied on an utterance is designed as follows. First, a pairwise relatedness score is computed between each pair of positions $(\mathbf{h}_i, \mathbf{h}_j)$ with a bilinear transformation, $c_{i,j} = \tanh(\mathbf{h}_i^\top \mathbf{W}^{(1)} \mathbf{h}_j)$, where $\mathbf{W}^{(1)} \in \mathbb{R}^{2h \times 2h}$ is a trainable parameter. Next, for each position \mathbf{h}_i , we convert its relatedness scores with all other positions into its overall relatedness scores with κ possible reasons using a linear transformation f ,

$$e_{i,k} = f(\{c_{i,j}\}_{j=1}^{|P| \text{ or } |Q|}) = \sum_{j=1}^{|P| \text{ or } |Q|} c_{i,j} \cdot W_{j,k}^{(2)} + b_k \quad (1)$$

where $\mathbf{W}^{(2)} = \{W_{j,k}^{(2)}\} \in \mathbb{R}^{|P| \text{ or } |Q| \times \kappa}$ and $\mathbf{b} \in \mathbb{R}^\kappa$ are trainable parameters. The philosophy behind Eq. 1 is that the relatedness distribution $\{c_{i,j}\}$ of an utterance implies segments in it that are internally compatible, which may correspond to different focal points (reasons). The transformation f then learns the mapping between those two. Finally, we obtain the attention weight $\mathbf{A}_{i,k}$ for each position i on each reason k by applying *softmax* over all $e_{*,k}$ s,

$$\mathbf{A}_{i,k} = \frac{\exp(e_{i,k})}{\sum_{j=1}^{|P| \text{ or } |Q|} \exp(e_{j,k})} \quad (2)$$

With \mathbf{A} obtained, we can compute an utterance's reason encoding as the sum of its RNN encoding $\{\mathbf{h}_i\}$ weighted by \mathbf{A} : $\mathbf{r}_k = \sum_{i=1} a_{k,i} \mathbf{h}_i$, where $\mathbf{r}_k \in \mathbb{R}^{2h}$ is the encoding for the k th reason. We use $\mathbf{R}^{P(Q)} = [\mathbf{r}_1^{P(Q)}, \dots, \mathbf{r}_\kappa^{P(Q)}] \in \mathbb{R}^{2h \times \kappa}$ to denote the *reason matrix* for the utterance $P(Q)$.

It is worth noting that the above self-attention mechanism in our reason encoding can also be seen as a variant of *multi-dimensional* self-attention, as we simultaneously learn multiple attention vectors for the different reasons implied in an utterance.

Stance Comparator: Now we compare the stances of P and Q based on their reason matrices. Since we have captured multiple reasons in each utterance, all the differences between their reasons must be considered. We thus take a *reason-wise* comparing approach, where every possible pair of reasons between P and Q is compared. We employ two widely used operations for the comparison, i.e., *multiplication*: $\mathbf{s}_{i,j}^{\text{mul}} = \mathbf{r}_i^P \odot \mathbf{r}_j^Q$ and *subtraction*: $\mathbf{s}_{i,j}^{\text{sub}} = (\mathbf{r}_i^P - \mathbf{r}_j^Q) \odot (\mathbf{r}_i^P + \mathbf{r}_j^Q)$, where \odot denote element-wise multiplication. We then

aggregate all the differences resulting from each operation into a single vector, by using a global max-pooling to signal the *largest* difference with respect to an operation,

$$\begin{aligned} \mathbf{s}^{\text{mul}} &= \text{global_max_pooling}(\{\mathbf{s}_{i,j}^{\text{mul}} | i, j \in [1, \kappa]\}) \\ \mathbf{s}^{\text{sub}} &= \text{global_max_pooling}(\{\mathbf{s}_{i,j}^{\text{sub}} | i, j \in [1, \kappa]\}) \end{aligned} \quad (3)$$

The concatenation of the two difference vectors $\mathbf{s} = [\mathbf{s}^{\text{mul}}; \mathbf{s}^{\text{sub}}]$ forms the output of this module.

(Dis)agreement Classifier: Finally, a classifier is deployed to produce the (dis)agreement class probability $\hat{\mathbf{y}} = \{\hat{y}_1, \hat{y}_2, \hat{y}_3\}$ based on the comparison result \mathbf{s} , which consists of a two-layer feed-forward network followed by a softmax layer, $\hat{\mathbf{y}} = \text{softmax}(\text{FeedForward}(\mathbf{s}))$.

Optimisation: To train our model, we use the multi-class cross-entropy loss,

$$\mathcal{L}(\theta) = - \sum_i^N \sum_j^3 y_j^{(i)} \log \hat{y}_j^{(i)} + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (4)$$

where N is the size of training set, $y \in \{Agree, Disagree, Neither\}$ is the ground-truth label indicator for each class, and \hat{y} is the predicted class probability. λ is the coefficient for L_2 -regularisation. Θ denotes the set of all trainable parameters in our model.

Minimising Eq. 4 encourages the comparison results between the extracted reasons from P and Q to be stance-predictive.

3 Related Work

Our work is mostly related to the task of detecting agreement and disagreement in online discussions. Recent studies have mainly focused on classifying (dis)agreement in dialogues (Abbott et al., 2011; Wang and Cardie, 2014; Misra and Walker, 2013; Allen et al., 2014). In these studies, various features (e.g., structural, linguistic) and/or specialised lexicons are proposed to recognise (dis)agreement in different dialogic scenarios. In contrast, we detect stance (dis)agreement between independent utterances where dialogic features are absent.

Stance classification has recently received much attention in the opinion mining community. Different approaches have been proposed to classify stances of individual utterances in ideological forums (Murakami and Raymond, 2010; Somasundaran and Wiebe, 2010; Gottopati et al., 2013; Qiu et al., 2015) and social media (Augenstein et al.,

2016; Du et al., 2017; Mohammad et al., 2017). In our work, we classify (dis)agreement relationships between a pair of stance-bearing utterances.

Reason information has been found useful in argumentation mining (Lippi and Torroni, 2016), where studies leverage stance and reason signals for various argumentation tasks (Hasan and Ng, 2014; Boltuzic and Snajder, 2014; Sobhani et al., 2015). We study how to exploit the reason information to better understand the stance, thus addressing a different task.

Our work is also related to the tasks on textual relationship inference, such as textual entailment (Bowman et al., 2015), paraphrase detection (Yin and Schütze, 2015), and question answering (Wang et al., 2016). Unlike the textual relationships addressed in those tasks, the relationships between utterances expressing stances do not necessarily contain any rephrasing or entailing semantics, but they do carry discourse signals (e.g., reasons) related to stance expressing.

4 Experiments

4.1 Setup

Dataset: The evaluation of our model requires a corpus of agreed/disagreed utterance pairs. For this, we adapted a popular corpus for stance detection, i.e., a collection of tweets expressing stances from SemEval-2016 Task 6. It contains tweets with stance labels (*Favour*, *Against*, and *None*) on five topics, i.e., *Climate Change is a Real Concern* (CC), *Hillary Clinton* (HC), *Feminist Movement* (FM), *Atheism* (AT), and *Legalization of Abortion* (LA). We generated utterance pairs by randomly sampling from those tweets as follows: *Agreement* samples: 20k pairs labelled as (*Favour*, *Favour*) or (*Against*, *Against*); *Disagreement* samples: 20k pairs as (*Favour*, *Against*), (*Favour*, *None*), or (*Against*, *None*); *Unknown* samples: 10k pairs as (*None*, *None*)¹.

Baselines: We compared our method with the following baselines: 1) BiLSTM: a base model for our task, where only the RNN encoder is used to encode the input; 2) DeAT (Parikh et al., 2016): a popular attention-based models for natural language inference. 3) BiMPM (Wang et al., 2017): a more recent natural language inference model

¹Fewer *unknown* pairs being sampled is due to the inherently fewer *none*-stance tweets in the original corpus.

Topic	BiLSTM	DeAT	BiMPM	RCN (Our)
CC	68.1±0.6	70.9±0.7	71.5±0.6	73.0±0.5*
HC	52.5±0.6	56.9±0.4	56.4±0.7	58.6±0.4**
FM	58.3±0.6	60.6±0.7	59.8±0.7	64.4±0.5**
AT	67.5±0.4	69.5±0.5	70.3±0.6	72.2±0.4*
LA	61.3±0.3	63.2±0.6	62.4±0.4	64.5±0.4**

Two tailed t-test: ** $p < 0.01$; * $p < 0.05$

Table 2: Classification performance of the compared methods on various topics, measured by the averaged macro F1-score over ten runs on the test data.

where two pieces of text are matched from multiple perspectives based on pooling and attention.

Training details: An 80%/10%/10% split was used for training, validation and test sets. All hyper-parameters were tuned on the validation set. The word embeddings were statically set with the 200-dimensional GloVe word vectors pre-trained on the 27B Twitter corpus. The hidden sizes of LSTM and FeedForward layers were set to 100. A light dropout (0.2) was applied to DeAT and heavy (0.8) to the rest. ADAM was used as the optimiser and learning rate was set to 10^{-4} . Early stopping was applied with the patience value set to 7.

4.2 Results

Table 2 shows the results of our method and all the baselines on tasks with different topics. We can first observe that the proposed RCN consistently outperformed all the baselines across all topics. Despite being modest, all the improvements of RCN over the baselines are statistically significant at $p < 0.05$ with a two-tailed t-test. Among these methods, BiLSTM performed the worst, showing that only using the RNN encoder for sequence encoding is not sufficient for obtaining optimal results. DeAT and BiMPM performed similarly well; both used attention to compare the utterances at a fine-grained level, resulting in a 2~5% boost over BiLSTM. Finally, RCN performed the best, with relative improvements from 2.1% to 10.4% over the second best. As all the compared methods shared the same RNN encoding layers, that RCN performed empirically the best demonstrates the efficacy of its unique reason encoder and stance comparator in boosting performance.

4.3 Analysis

In this section, we study what has been learned in the reason encoder of RCN. In particular, we show the attentive activations in the reason en-

ID	Label	Topic	Tweet 1	Tweet 2
1	Agree	HC (A, A)	Reason 1: @HillaryClinton is a liar & corrupt . Period. End of story.	Reason 1: @HillaryClinton lies just for the fun of it, its CRAZY!!!!!!
2	Disagree	LA (F, A)	Reason 1: I would never expect an 11 year old girl to have to carry a pregnancy to term	Reason 1: Actually, child murder is far worse these days. We live in more savage times.
3	Agree	CC (F, F)	Reason 1: Living an unexamined #life may be easier but leads to disastrous consequences. Reason 2: Living an unexamined #life may be easier but leads to disastrous consequences .	Reason 1: There's no more normal rains anymore. Always storms, heavy and flooding. Reason 2: There's no more normal rains anymore. Always storms , heavy and flooding.

Table 3: The heatmaps of the attention weights assigned by the attention layer in the reason encoder to three tweet-pair examples. In each example, we show the text of each tweet, the topic, the correct (dis)agreement label, and the stance of each tweet (F: Favour, A: Against).

Topic	Top reason words ranked by attention weights
CC	environment, climate, sustainability, safety, economy, community, good, kill, drought, insane, proud, co2, coal, clean, green
HC	candidate, freedom, liberal, disappointed, greed, democrat, cheat, illegal, best, economy, war, american, republican, cutest
FM	women, husband, divorce, girlfriend, adorable, ignorant, rights, behaved, marriage, infanticide, gender, queen, child, equality
AT	fear, evil, jesus, human, truest, god, pray, belief, religion, ancient, tribulation, love, sovereign, church, secular, ignorance
LA	pregnant, abortionist, murder, accidental, right, fertility, justice, illegal, democrat, marriage, government, motherhood, freedom

Table 4: The reason words learned on various topics.

coder (i.e., **A** in Eq. 2), and see if reason-related contents could draw more attention from RCN.

Visualising attention signals in tweets: Table 3 shows the attention activations on three examples of tweet pairs chosen from our test set. For the first two, we set the number of reasons to be attended to as one. It can be seen that the parts of the tweets that received large attention weights (the highlighted words in Table 3) were quite relevant to the respective topics; *liar*, *corrupt*, and *lie* are words appearing in news about *Hillary Clinton*; *girl*, *pregnancy*, and *murder* are common words in the text about *Legalisation of Abortion*. Also, most of the highlighted words have concrete meanings and are useful to understand why the stances were taken. The last row shows a case when two reasons had been attended to. We observe a similar trend as before that the highlighted contents were topic-specific and stance-revealing. Moreover, since one more reason dimension was added to be inferred in this case, RCN was able to focus on different parts of a tweet corresponding to the two reasons.

Visualising learned reasons: We also visualised the reasons learned by our model, represented as the words assigned with the largest attention weights in our results (i.e., 1.0). Table 4 shows samples of such reason words. We

see that the reason words have strong correlations with the respective topics, and, more importantly, they reflect different reason aspects regarding a topic, such as *economy* vs. *community* on *Climate Change is a Real Concern* and *culture* vs. *justice* on *Legalisation of Abortion*.

In summary, both the visualisations in Table 3 and 4 show that the attention mechanism employed by RCN is effective in finding different reason aspects that contribute to stance comparison.

5 Conclusion and Future Work

In this paper, we identify (dis)agreement between stances expressed in paired utterances. We exploit the reasons behind the stances and propose a reason comparing network (RCN) to capture the reason information to infer the stance (dis)agreement. A quantitative analysis shows the effectiveness of RCN in recognising stance (dis)agreement on various topics. A visualisation analysis further illustrates the ability of RCN to discover useful reason aspects for the stance comparison.

In the future, this work can be progressed in several ways. First, it is necessary to evaluate our model on more stance data with different linguistic properties (e.g., the much longer and richer stance utterances in posts or articles). Second, it is important to show how the learned embedded reasons can help downstream applications such as stance detection. Finally, it would be insightful to further visualise the reasons in the embedded space with more advanced visualisation tools.

Acknowledgements

We thank all anonymous reviewers for their constructive comments. We would also like to thank Xiang Dai for his helpful comments on drafts of this paper.

References

- Rob Abbott, Marilyn Walker, Pranav Anand, Jean E Fox Tree, Robeson Bowmani, and Joseph King. 2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, pages 2–11.
- Amjad Abu-Jbara, Mona Diab, Pradeep Dasigi, and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 399–409.
- Kelsey Allen, Giuseppe Carenini, and Raymond Ng. 2014. Detecting disagreement in conversations using pseudo-monologic rhetorical structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1169–1180.
- Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885.
- Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. Stance classification of context-dependent claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 251–261.
- Filip Boltuzic and Jan Snajder. 2014. Back up your stance: Recognizing arguments in online discussions. In *ArgMining@ACL*, pages 49–58.
- Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642.
- Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a” siamese” time delay neural network. In *NIPS*, pages 737–744.
- Jianpeng Cheng, Li Dong, and Mirella Lapata. 2016. Long short-term memory-networks for machine reading. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 551–561.
- Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 539–546.
- Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 3988–3994.
- Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2016. Weakly supervised tweet stance classification by relational bootstrapping. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1012–1017.
- William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 1163–1168.
- Swapna Gottopati, Minghui Qiu, Yanchuan Sim, Jing Jiang, and Noah Smith. 2013. Learning topics and positions from debatepedia. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1858–1868.
- Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356.
- Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762.
- Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP)*, pages 59–70.
- Baotian Hu, Zhaopeng Tu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2015. Context-dependent translation selection using convolutional neural network. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, pages 536–541.
- Marco Lippi and Paolo Torrioni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2):10.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119.
- Amita Misra and Marilyn Walker. 2013. Topic independent identification of agreement and disagreement in social media dialogue. *Proceedings of the SIGDIAL 2013 Conference*, pages 41–50.

- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *SemEval@ NAACL-HLT*, pages 31–41.
- Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26.
- Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan, and Zhi Jin. 2016. Natural language inference by tree-based convolution and heuristic matching. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 130–136.
- Akiko Murakami and Rudy Raymond. 2010. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 869–875.
- Ankur P Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. 2016. A decomposable attention model for natural language inference. *arXiv preprint arXiv:1606.01933*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*, pages 1532–1543.
- Minghui Qiu, Yanchuan Sim, Noah A Smith, and Jing Jiang. 2015. Modeling user arguments, interactions, and attributes for stance prediction in online debate forums. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 855–863.
- Sara Rosenthal and Kathy McKeown. 2015. I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 168–177.
- Parinaz Sobhani, Diana Inkpen, and Stan Matwin. 2015. From argumentation mining to stance classification. In *ArgMining@HLT-NAACL*, pages 67–77.
- Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124.
- Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 116–125.
- Dhanya Sridhar, Lise Getoor, and Marilyn Walker. 2014. Collective stance classification of posts in online debate forums. In *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pages 109–117.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, pages 5998–6008.
- Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 592–596.
- Bingning Wang, Kang Liu, and Jun Zhao. 2016. Inner attention based recurrent neural networks for answer selection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1288–1297.
- Lu Wang and Claire Cardie. 2014. Improving agreement and disagreement identification in online discussions with a socially-tuned sentiment lexicon. *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 97–106.
- Wen Wang, Sibel Yaman, Kristin Precoda, Colleen Richey, and Geoffrey Raymond. 2011. Detection of agreement and disagreement in broadcast conversations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 374–378.
- Zhiguo Wang, Wael Hamza, and Radu Florian. 2017. Bilateral multi-perspective matching for natural language sentences. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 4144–4150.
- Wenpeng Yin and Hinrich Schütze. 2015. Convolutional neural network for paraphrase identification. In *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, pages 901–911.